CONSER Research Project: Second Preliminary Report
4/27/2005

Background

On Jan. 28, 2005, Robert Bremer created a snapshot of the CONSER database. Using that snapshot, he prepared two randomly-selected sets of 1,000 records. At the same time, he created a summary report of 856s in the CONSER database:

- Total records with at least one 856 field: 50,176
- Total URLs: 71,809
-     Records with more than 1 URL: 13,525
-     Records with more than 5 URLs: 85
-     Average URLs per record: 1.43
-     Total GPO PURLs: 7,999
-     Total BibPURLs: 3,129
-     Total URLs with ".gov":   10,420
-     Total URLs with ".org" or ".edu":   28,069
-     Total URLs with ".com":   22,290
-     Total URLs with "?genre":   302
- Representing single-record approach: 28,619
- Representing separate-record approach:21,557

Records in the first sample represent the "single record" approach to providing access to online serials. They were selected according to the following criteria:
- CONSER record (has 042)
- for print (Form="blank")
- with at least one 856 field.

Records in the second sample represent the "separate record" approach; they met the following selection criteria:
- CONSER record (has 042)
- for electronic version (Form="s")

The resulting file of 2000 randomly-selected OCLC numbers was delivered to the OpenURI Task Group on Feb. 24, 2005.

Since that time, Luiz Mendes (UCLA) has been analyzing 800 of the records (400 records from each sample). On Apr. 26th, he completed an evaluation of 400 records describing print serials and 195 records describing remote access serials[1] and coded his analysis in an Excel spreadsheet. As part of the evaluation process, he also reviewed 203 corresponding online records for the print records and 131 corresponding print records for the online records. In all, the total number of records considered in this preliminary evaluation is 929.

In early April, Luiz Mendes prepared a preliminary report covering 240 records from the samples. His observations and conclusions have been incorporated (in part) into this report.

Overview

This overview is divided into two parts. The first part discusses evaluation results for records for print serials; the second part, evaluation results for the sample of aggregator neutral records. Within each part are segments providing: a general summary and specific counts for paid-access serials, government documents, and other freely-available serials.

---

[1] For brevity, the discussions below use the term "print records" for the former and "online records" for the latter

A note of caution: All of the statistics are subject to change, since the analysis is still underway. Analysis for the sample of print serial records is close to completion, but additional records need to be added to compensate for those that could not be used (incorrectly coded or lacking sufficient data). Analysis of the sample of online serial records is about half-done; the statistics for that part are preliminary.

Print Records

Print Records: General Summary

Of the 400 records, 4 had to be removed (incorrectly coded as print records or uncategorizable). The remaining 396 records fell into the following general categories:
- 188 records with "paid" online access[1]
- 98 records for government documents
- 113 records for other freely-available serials

[Note: The categories total 399 because 3 of the serials had both paid and free access]

Within the 856 field, subfield $u links in these records (excluding inactive links) predominantly included URLs. Out of 561 URIs, a breakdown by types shows:
- 480 URLs
- 40 PURLs
- 37 GPO PURLs
- 4 URLs with OpenURL syntax

Print Records: Paid Access

"Paid access e-serials" is used in this document as a short form for serials that are licensed or available online through subscription. The sample included 188 records for paid resources (=47.12% of the sample). Of these records, 38 had invalid URIs; that is, 20.21% of the paid access records had an invalid URL. The invalid URIs were categorized as follows:
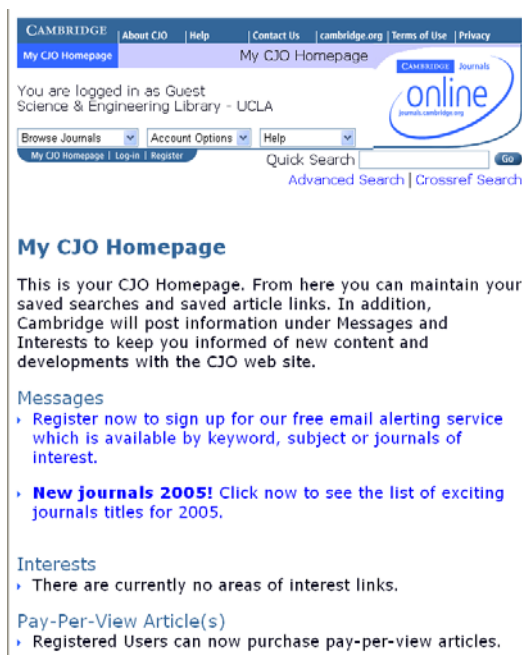- 37 URLs
- 1 URL with OpenURL syntax

As part of the analysis, the project also included a review of the aggregator neutral records corresponding to the 188 print records. 143 print records (76%) had a corresponding record for the online version. The percentage of record pairs is higher for paid access records than for any other category in the sample (i.e., much higher than for government documents or freely available e-serials).

Comparing print/aggregator neutral record pairs, several other patterns emerge. Most obvious are differences between the number of URIs recorded, which e-products are represented, and the place pointed to by the URL within the web site (when both records have URIs for the same e-product).

The OCLC records for Journal of anatomy illustrate a difference in URL for the same e-productr: the print record links to the journal home page; the online record, to a general information page.

2246129	journal home page	http://www.journals.cambridge.org/jid%5FANA
41883951 (online)	Information page for publisher site	http://www.journals.cup.org/

One might expect the record for the online version to contain the same number or more links than are present on the print record. However, this is not always the case. For example, the online record for Economic policy links only for OCLC FirstSearch. The print record also includes links for Ingenta Connect and Blackwell Synergy.

| 13365818 | print | http://firstsearch.oclc.org
http://firstsearch.oclc.org/journal=0266-4658;screen=info;ECOIP
http://www.ingenta.com/journals/browse/bpl/ecop
http://www.blackwell-synergy.com/rd.asp?goto=journal&code=ecop |
| 40337313 | online | http://firstsearch.oclc.org
http://firstsearch.oclc.org/journal=0266-4658;screen=info;ECOIP |

Similarly, New York Jewish week has two links in the print record and no links in the record for the online version.

| 9144716 | print | http://www.softlineweb.com/softlineweb/ethnic.htm

http://bibpurl.oclc.org/web/1830 ‡u http://www.thejewishweek.com |
| 44893552 | online | No 856 |

Print records for paid serials include 341 URIs (about 1.8 per record):
- 336 URLs
- 4 URLs with OpenURL syntax
- 1 PURL (OCLC PURLs, not BibPURLs)

While URLs with OpenURL syntax seem to be becoming more common in records, the increased use of this URI category was not been reflected in the sample. As outlined above, only four instances appeared in the sample. Two examples of this category are:

OCLC *1794709
Latin American perspectives
http://openurl.ingenta.com/content?genre=journal&issn=0094-582X

OCLC *28906651
Multimedia systems.
http://www.springerlink.com/openurl.asp?genre=journal&issn=0942-4962

Print Records: Government Documents

The category of government documents consists of records for serials issued by government agencies at any level: federal, state, local, or international. The sample included 98 records for government documents (=24.56% of the sample). Within this category, 19 records had invalid URIs (19.39%). The breakdown for invalid URIs was:

- 16 URLs
- 2 GPO PURLs
- 1 BibPURL

Print records for government document serials include a total of 107 URIs, about 1 per record. Not surprisingly, the records often had URL substitutes:

- 37 GPO PURLs
- 20 PURLs (state documents)
- 50 URLs

Twenty-eight (28.57%) government document records had corresponding aggregator neutral records. As with paid access serials, paired print/online records for non-federal government documents records also demonstrate some of the asymmetry found for paid resources. The sample included an extreme case, State of Hawaii data book, whose print version has one link and whose online record has no links at all. The link is invalid—leads to a redirect page for the site. (Note: the serial still exists, the web site was just re-: http://www3.hawaii.gov/dbedt/index.cfm?section=statistics_and_economic_information516)

1783344	print	http://www.state.hi.us/dbedt/
51734400	online	

Record pairs are less common with federal documents than with paid access serials, but occur; some with unequal links. One such case is Turkeys raised (Washington, D.C.). Here, the online record has GPO PURLs, while the print record has a URL.

46926076	Print	http://usda.mannlib.cornell.edu/reports/nassr/poultry/pth-bbt/
52497082	Online	http://purl.access.gpo.gov/GPO/LPS32772 http://purl.access.gpo.gov/GPO/LPS4563 (email address)

A more difficult problem with 856 links in government documents relates to agencies' web practices. Unlike licensed resources, e-serials on government agency web sites are often accessible only indirectly. For example, the NIJ research portfolio (OCLC# 43314515) is available through the National Institute of Justice web site (http://www.ojp.gov/nij/pubs.htm). But once a user gets to the top-level web site, what then?

In contrast, the New York State Library provides links to its repository issues. For example, I love New York travel guide (OCLC# 4238382) has a link to: http://purl.org/net/nysl/nysdocs/4238382. The PURL leads to a page listing issues available through the repository. While the issues list page omits the contextual flavor of the original web site, the goal of getting the serial is better met.

Print Records: Other Freely-Available Online Serials

For the purpose of this discussion, "freely-available e-serials" excludes online serials issued by a government agency. Examples of freely-available e-serials include organization newsletters, bulletins issued by educational institutions, and some open-access e-journals. The sample contains 113 records with links to

freely-available serials (=28.32% of the sample). Of the 113 records, 25 (22.12%) have invalid URLs. This was the highest rate of invalid URLs among the three categories of print records. Three PURLs and 22 URLs are invalid.

Thirteen of the records for freely-available serials (11.5%) have corresponding records for the online version. The freely-available e-serial category clearly showed the strongest tendency toward the single-record approach. Nine of the pairs have consistent 856 URIs for both the print and online record; four diverged. One example of divergence in a record pair for a freely-available serial is *Koreana*. *Koreana* has a PURL in the print record & no link in the online record.

17220991	Print	http://bibpurl.oclc.org/web/1580 ‡u http://www.kf.or.kr/koreana/
46849255	Online (NonCONSER)	No link

Print records for freely-available serials (other than government documents) included a total of 113 URIs (1 URI per record) :
- 97 URLs
- 19 PURLs

Aggregator Neutral Records

Aggregator Neutral Records: General Summary

So far, 196 records in the sample have been analyzed. Two records were rejected that represented CD-ROMs rather than remote access serials, leaving 194 records. These records fall into three roughly equal groups:
- 70 Paid access
- 64 Government documents
- 62 Freely-available (other than government documents)

It is difficult to predict whether this distribution will continue to hold true as the rest of the sample is analyzed. But the pattern is an intriguing contrast to the sample for print records, where over 47% were records for paid access serials.

Within the sample, 47 records had at least one invalid URL, an average of 24.2%.

The records contain 300 URIs, of the following types:
- 257 URLs
- 28 GPO PURLs
- 11 PURLs
- 3 Institution-specific URLs
- 1 Mailto

Aggregator Neutral Records: Paid Access Serials

Among the 70 records for paid access serials, 67 have corresponding print records (=95.71%). Seven records contain at least one invalid URL (=10%). Altogether, the paid access records include 138 URIs, an average of 1.97 URI per record:
- 135 URLs
- 3 Institution-specific URLs

Aggregator Neutral Records: Government Documents

The 65 records for government documents have corresponding print records in 32 cases (=49.2% of the time). Twenty-four of the records include at least one invalid URI; of these:

- 23 URLs
- 1 GPO PURL

The records contain a total of 92 URIs. Put another way, each record has an average of 1.4 URIs. The URIs may be further categorized:

- 64 URLs
- 28 GPO PURLs

Aggregator Records: Other Freely-Available Serials

Sixty-two freely-available serials (other than government documents) have been identified, 28 of which have corresponding print records (=45.31%). 15 of the records have at least one invalid URI (=24.19%). This very high error rate may or may not be significant; it will be interesting to monitor whether this percentage changes once all of the records have been tabulated.

In all, the records contain 69 links to the online version (about 1 URI per record):

- 57 URLs
- 11 PURLs
- 1 Mailto

Discussion Issues for CONSER Members

Many issues surfaced during this project; several are listed below. A more complete list will accompany the final report.

- Form of issuance: Sometimes, the link in a record leads to a resource which may have the same content in a different form of issuance—an integrating resource rather than a serial. Should "856 41" be used to include links for integrating resources, or should these fields only be used to link to e-serials?

- Choice of URI target page: Should a list of preferences be elaborated for choosing which particular page of an online serial to point to with the URI? Where options exist, what is the preferred place for a URI to point to?: Journal issues list? Journal contents page? Journal home page?

- Maintenance of 776: What commitment should catalogers have to identifying print/online pairs through 776 fields? When completing maintenance work (e.g., descriptive change) on print records that have 856 links, should CONSER catalogers check for record for online serial & add 776 fields?

- Maintenance of 856: The number of records in the print sample showed a high percentage of records with invalid URLs, across all categories. Should CONSER records continue to include URLs on bib records?
  - If yes (if CONSER continues to include these links), in the case of print/online record pairs should BOTH the print & Aggregator neutral record have these links? Or, should all links be moved to the Aggregator neutral record?

  - If CONSER records continue to include these links, should CONSER catalogers remove invalid links or continue to add $z comments?

  - If both print & online records continue to include these links, what commitment should CONSER catalogers have toward link maintenance? What commitment should CONSER catalogers have toward maintaining links on print/online in tandem?

  - If 856 URI links continue to be included in records, are there any ways to represent the links that will reduce the problem of maintenance within the CONSER record (e.g., OpenURI, "info" URI)?